ARTICLE

Check for updates

Exploring political orientation: a comparative analysis of news media and government responses to the Russia-Ukraine conflict

Seongkyu Ko[1,2], Junyeop Cha[1,3✉], Eunil Park[1,3✉] & Angel P. del Pobil[3]

On February 24, 2022, the Russian government launched a "special military operation," triggering the Ukraine-Russia conflict. In response, the United States government condemned Russia's actions and continued to provide economic and military support to Ukraine. However, within U.S. society, policy perspectives on Ukraine varied depending on political orientation. To analyze these differences, we collected news articles produced by U.S. media outlets and examined the distinctions between conservative and progressive sources. For the analysis, we used a word-frequency-based psycholinguistic analysis tool to construct seven word categories. Subsequently, we developed a classification model based on the collected articles and applied it to categorize official government documents according to their political orientation. The model classified 291 documents, identifying 64 as conservative and 227 as progressive. Additionally, we employed topic modeling to measure the similarity between conservative and progressive media. Using Jensen-Shannon Divergence (JSD), we found that government documents exhibited a greater similarity to conservative media sources. Furthermore, we analyzed the implicit intentions embedded within official government documents. This study demonstrates the feasibility of using computational methods to measure political orientation in texts produced by news media and government sources regarding the Russia-Ukraine conflict.

[1] Sungkyunkwan University, Seoul, Republic of Korea. [2] Ministry of National Defense, Seoul, Republic of Korea. [3] Jaume I University, Castelló de la Plana, Spain. ✉email: jycha95@g.skku.edu; eunilpark@skku.edu

Introduction

On February 24, 2022, the Russian government ordered the initiation of a special military operation, leading to an outbreak of the Ukraine-Russia conflict. Although several experts presented indispensable outcomes of the conflict, it is important to note that the current situation between Russia and Ukraine is not merely a short-term outcome (Brunk and Hakimi, 2022). The Russian government claimed that Ukraine fell into the hands of extremists after the dethronement of pro-Russian former Ukrainian president, Viktor Yanukovych, in 2014. As a result, the Russian government took action to annex Crimea (Kudelia, 2022). In late 2021, the Russian government began building up military forces near the Ukrainian border, escalating tension (Kirby, 2022) due to Ukraine's attempts to join the North Atlantic Treaty Organization (NATO). Consequently, the Russian government launched an invasion of Ukraine (Vajriyati et al. 2022)

Following the invasion, the United States condemned Russia's actions (Whitehouse, 2022b) and garnered international support for imposing sanctions against Russia (Whitehouse, 2022a). Economic pressure was exerted on Russia's energy, mining, finance, defence, and industrial sectors, with export and import restrictions imposed on Russia and its supportive countries. On the military front, the U. S. provided weapons with related systems and supported equipment to help Ukraine's security directly or indirectly (Whitehouse, 2023). However, depending on political orientations, there have been doubts in the United States about the appropriateness of aiding Ukraine (Pew Research Center, 2023).

In fact, Pew Research Center (2023) conducted a survey in January 2023, revealing that Republicans perceived aid to Ukraine as excessive compared to Democrats. This survey highlighted how viewpoints can differ based on the political orientation within a nation. This is the central proposition of our research.

Traditionally, various news media outlets, including CNN and FOX, engage in ongoing discourse and reflect their own stances on specific issues (Blanchard, 2013), with each news media having its own political orientation (Entman, 2007).

Theoretical frameworks, such as Framing Theory (Entman, 1993) and Moral Foundations Theory (Haidt and Graham, 2007), provide a deeper understanding of why media outlets exhibit distinct political orientations through their choice of language. Framing Theory, proposed by Entman (1993), posits that media organizations selectively emphasize certain aspects of an issue while ignoring others, effectively shaping public perception and interpretations. Different news outlets, influenced by their ideological perspectives, may present the same event through varying frames, thus embedding their political leanings in their articles.

Additionally, Haidt's Moral Foundations Theory (Haidt and Graham, 2007) suggests that individuals and groups differ systematically in their moral values, which influences political ideology and discourse. According to this theory, conservative media may highlight moral dimensions, such as loyalty, authority, and sanctity, whereas progressive media may emphasize care, fairness, and social justice. Such moral framing inherently shapes the language used in media coverage of politically contentious issues, such as the Russia-Ukraine conflict.

Based on this observation, we assumed that articles produced by the news media might indicate their political leanings, particularly concerning the Russia-Ukraine conflict.

Consequently, we formulated the following Research Questions (RQs) to investigate this assumption

- **RQ1: Do the articles produced by the news media on the issue contain their own political orientations?**

Governments inherently possess their own political orientations (Potrafke, 2018). Presidents typically have prior political experience within a party, which informs their perspectives and decision-making framework (Walker, 2009). Consequently, the administration led by the president formulates policies and implements decisions based on these orientations. Based on this premise, we define the second research question:

- **RQ2: To what extent do government documents announcing and implementing specific policies reflect political ideology on the issue?**

We used two methodologies to address these research questions. Initially, we employed sentiment analysis, which is a computational approach that assesses opinions, emotions, and subjectivity in text (Medhat et al., 2014). Building on previous studies on sentiment analysis, we refined our methodology to explore whether articles mirror the political leanings of news media. Furthermore, to adopt a computational approach, we integrated artificial intelligence models. Artificial intelligence encompasses technologies such as machine learning and deep learning, enabling systems to accurately interpret external data, learn from it, and flexibly adapt to achieve specific goals and tasks (Kaplan and Haenlein, 2019). In this study, we incorporated machine learning and deep-learning models to perform computational tasks. By training these models with articles concerning the Russia-Ukraine conflict from various U.S. media outlets, we conducted experiments to evaluate their effectiveness in detecting the political orientation embedded within news articles, achieving an accuracy rate of 94.88%.

We designed the following study to address these two research questions:

First, we applied statistical methodologies to examine whether there were significant differences among the articles produced by media outlets with distinct political orientations. Specifically, we employed a psycholinguistic analysis. Psycholinguistic analysis investigates the relationship between language and cognitive processes by focusing on how linguistic structures influence perception, memory, and thought. This field explores how language shapes cognitive function, and in turn, how these functions affect language use (Altmann, 1998). Based on previous studies, we refined our methodology to assess whether news articles reflected the political orientation of their respective media sources. Second, we employed natural language processing (NLP) models to classify media articles and evaluate their accuracy in capturing political orientations. Artificial intelligence (AI) encompasses technologies such as machine learning and deep learning, enabling systems to accurately interpret external data, learn from it, and flexibly adapt to achieve specific tasks and objectives (Kaplan and Haenlein, 2019). In this study, we integrated machine-learning and deep-learning models to conduct computational analyses. Third, to determine whether government documents contain implicit political intentions, we applied the aforementioned statistical methodologies and AI models for verification. Finally, we applied topic modeling to uncover hidden themes in the documents for a more in-depth analysis.

Literature review

Several scholars explored textual political differences. For instance, (Kangas, 2014) conducted an experiment during the 2008 presidential election and analyzed the relationship between discourse and the personalities of presidential and vice-presidential candidates in the United States. They collected 144 speech samples and employed the Linguistic Inquiry and Word Count (LIWC) dictionaries to apply one-way analysis of

variance (ANOVA). Based on these results, linguistic differences between candidates were observed.

In the 2016 presidential election, Abe (2018) investigated the underlying reasons for the strong association between evaluations of candidate personalities and voting preferences. They collected a dataset from 160 participants and utilized four LIWC-based customization dictionaries based on Haidt's moral foundation model (Graham et al., 2009), and four customized dictionaries based on Schwartz's basic human values model (Jones et al., 2018). The results revealed that Trump supporters scored significantly higher on measures of authenticity, collectivist pronouns, and basic human conservative values. They also scored significantly lower on moral foundation liberal values than Clinton supporters. Interestingly, the two groups did not show significant differences in their use of positive emotion words, but Clinton supporters used significantly higher levels of negative emotion words than Trump supporters did.

In the 2020 presidential election, Körner et al. (2022) explored the linguistic differences between Donald Trump and Joe Biden using 500 speeches and 15,000 tweets from the U.S. presidential election. They focused on their distinct forms of power and analyzed data from speech and tweets using the LIWC categories. They found that Biden's linguistic style was analytical, whereas Trump's had a positive emotional tone. Moreover, they indicated that Biden's tweets contained more social- and virtue-related words.

In addition, several scholars conducted a series of sentimental analyses in the political field. For instance, Rodriguez (2020a) applied the moral politics theory proposed by Lakoff (2010) to analyze the retrospectives of Republican and Democratic presidential candidates' campaigns between 2000 and 2016. Külz et al. (2023) analyzed the evolution of the linguistic tone of the U.S. politicians reported in online media between 2008 and 2020. Sterling et al. (2020) investigated the spontaneous, naturally occurring use of language to observe significant differences in the linguistic styles of liberals and conservatives. Lee et al. (2023) aimed to investigate the presence of media bias in nonpolitical articles. To this end, the researcher examined articles from the two most biased news outlets. Through a human evaluation process, scholars measured and compared the differences between two groups of articles.

Related to RQ2, several prior studies attempted to classify whether specific news outlets and articles have political orientations. Park et al. (2011) identified the commentator's pattern and predicted the political orientation of the news articles. Over the course of one year, they collected 4780 news articles from Naver's political news section. Using Term Frequency-Inverse Document Frequency (TF-IDF) representation and a Support Vector Machine (SVM) classifier, they built a classification model that achieved 80% accuracy. Moreover, (Devatine et al., 2022) conducted a similar approach using 27,146 news articles, and employed deep-learning models such as Long Short-Term Memory (LSTM) to predict political orientation. Conover et al. (2011) attempted to predict the political orientation of Twitter users based on their political communication content and structure, specifically before the 2010 U.S. midterm elections. They used annotated data from 1000 users based on their political orientation, and trained an SVM model using hashtag metadata. Additionally, they applied latent semantic analysis to the content of user tweets, achieving an accuracy of 95%.

Several scholars also attempted to address tasks related to political orientation by considering the stance of governments (Alturayeif et al., 2023; Biesbroek et al., 2020). However, only a few studies specifically examined officially introduced documents of governments to assess their political orientation. On the other hand, this study is distinct in that it analyzes hidden political biases in official U.S. government documents using a powerful transformer-based deep-learning model. In particular, by attempting a novel analysis that connects news and government documents, it expands the boundaries between political linguistics and policy analysis. Therefore, in line with this gap in existing research, our aim was to fill this gap by examining our RQs.

Methodology

Data collection. We used the Selenium library in Python for data collection and collected two types of data: articles produced by news media and official documents released by the U.S. government. Before data collection, four criteria were established. First, we collected data from February 24, 2022, to February 24, 2023, after the Russia-Ukraine conflict broke out. Second, we determined the political orientation of news media based on information from the Pew Research Center[1] and Media Bias Fact Check[2]. These organizations employ highly qualified journalists who follow carefully designed guidelines to make informed judgments (Baly et al., 2020). Third, we collected official documents of the U.S. government from press releases on the official websites of the White House and three administrative departments highly relevant to the Russia-Ukraine conflict: the Department of State (State), the Department of Defense (Defense), and the Department of Treasury (Treasury).

Finally, for data collection, we searched for the keywords "Ukraine," "Russia," "Putin," or "Zelensky." We specifically focused on gathering information related to the Russia-Ukraine war, such as real-time updates on the war situation, the United States' economic sanctions against Russia, or U.S. economic support for Ukraine. Only content directly associated with the specified keywords was included, while data unrelated to the Russia-Ukraine war was excluded.

Lastly, we searched for the keywords "Ukraine," "Russia," "Putin," or "Zelensky" and only selected data related to Russia-Ukraine conflict such as economic sanctions or U.S. economic aid to Ukraine, while data unrelated to the Russia-Ukraine confict was excluded.

Ultimately, 8104 articles were collected from various media outlets. Among them, 4789 articles were collected from four conservative news sources and 3315 articles were collected from nine progressive news sources. To ensure a balance between ideological perspectives, we applied random undersampling to the conservative news dataset, resulting in an equal number of articles from both the conservative and progressive sources. In addition, we collected 291 official documents from the U.S. government: 119 from the White House, 67 from the Treasury, 65 from the Defense, and the remaining from the State. The detailed data descriptions are presented in Table 1.

Psycholinguistic analyses. In this study, we conducted a comparative analysis to identify the characteristics of progressive and conservative articles through psycholinguistic analyses. To accomplish this, we utilized the Linguistic Inquiry and Word Count (LIWC) method. LIWC is a transparent analysis program that categorizes words into psychologically meaningful categories including word count, summary language variables, linguistic dimensions, other grammar, and psychological processes (Tausczik and Pennebaker, 2010). With the help of LIWC, researchers can count the number of sentiment words in a text, aiding in determining the overall sentiment level of the text (Pennebaker et al., 2015).

Additionally, we created LIWC categories based on previous studies that compared textual differences according to political orientation. Using these categories as references, we conducted a

Table 1 Summary of the collected dataset.

Category	Political orientation	Data source	Quantity	URL
Articles	Conservative	*Fox*	1891	https://www.foxnews.com/
		Wall Street Journal	1202	https://www.wsj.com/
		Forbes	1061	https://www.forbes.com/
		Breitbart	95	https://www.breitbart.com/
	Progressive	*New York Times*	854	https://www.nytimes.com/
		Politico	678	https://www.politico.com/
		PBS	493	https://www.pbs.com/
		The Guardian	463	https://www.theguardian.com/
		CNN	311	https://www.edition.cnn.com/
		MSNBC	298	https://www.msnbc.com/
		NPR	239	https://www.npr.com/
		The Washington Post	163	https://www.washingtonpost.com/
		Vox	73	https://www.vox.com/
	Total		8104	
Official documents	–	*White House*	119	https://www.whitehouse.gov/
		Department of Treasury	67	https://www.home.treasury.gov/
		Department of Defence	65	https://www.defence.gov/
		Department of State	40	https://www.state.gov/
	Total		291	

comparative analysis to discern the distinctions between progressive and conservative articles in terms of sentiment.

Classification

Experiment setting. In this study, we conducted experiments using Google Colab, which offers a cloud-based Jupyter notebook. This platform provides free access to GPUs and TPUs, enabling the efficient training and evaluation of machine learning models (Bisong and Bisong, 2019). The experiments used Python version 3.10.12, and the PyTorch framework was employed to construct the BERT model.

Preprocessing. To apply articles to the model, we converted the text into a numerical format. First, tokenization was applied to break down the text into individual words. We then removed stop words to exclude common words that did not carry significant meaning. We then converted the text to lowercase to ensure uniformity in the data. Additionally, we used lemmatization to transform words into their base or root forms, thereby reducing inflections and variations. In the final stage, we addressed the challenge of balancing the data classes. Class imbalance issues, common in various classification tasks, often cause a significant bias toward larger classes, leading classifiers to frequently overlook smaller classes (Ko et al., 2023). As our research dealt with imbalanced data classes, we executed a balancing task using a random sampling method to address this issue.

Classifier. Once the text was preprocessed, it was transformed into a numerical format suitable for machine learning and deep-learning models. In our study, we utilized two different architectures: eXtreme Gradient Boosting (XGB) and Bidirectional Encoder Representations from Transformers (BERT). These models allowed us to analyze and predict the characteristics and sentiments of the articles effectively.

- **XGB** is a widely used ensemble algorithm to achieve state-of-the-art results (Chen and Guestrin, 2016). Boosting is an ensemble technique in which new models are added to adjust the errors generated by existing models (Ogunleye and Wang, 2019).
- **BERT** is first introduced by Devlin et al. (2018). It uses a bidirectional transformer network for pre-training on a large corpus and fine-tuning specific tasks. The bert-base-

uncased, which is pre-trained BERT, is single layer feed-forward and the encoder is pre-trained with case-insensitive English text (Kim and Johnson, 2022).

The XGB model-IDF transformation method was used as the representation technique for the XGB model. On the other hand, for the BERT model, we utilized BERT embeddings as the representation.

Evaluation metric. To assess the performance of each model, we considered four vital performance metrics: accuracy, precision, recall, F-score, and ROC-AUC. These metrics are calculated based on a confusion matrix, which is a table that compares predicted and actual values (Hossin and Sulaiman, 2015).

Topic modeling. Topic modeling is a powerful tool capable of reading and understanding texts (Kherwa and Bansal, 2020). It is widely used as a statistical technique for extracting latent variables from large datasets (Agrawal et al., 2018). By analyzing word co-occurrence patterns within a collection of documents, topic modeling can uncover hidden topics (Wang et al., 2017). In this study, a pretrained BERTopic model (all-MiniLM-L6-v2) was employed, which is a neural topic modeling technique designed to extract coherent topics from text data. It utilizes pretrained transformer models to generate document embeddings that capture semantic nuances. These embeddings are then clustered, and a class-based TF-IDF is applied to generate topic representations (Grootendorst, 2022).

Experiment

Comparing two groups of media articles. The relationship between language and politics is reciprocal because it is closely interconnected (Chen and Madiyeva, 2022). Consequently, linguistic characteristics often reflect political orientations (Geis, 2012). Previous research extensively explored the connection between political orientations and linguistic features. Summarizing these studies, conservatives tend to employ linguistic features such as family-related language (Preoţiuc-Pietro et al., 2017; Rodriguez, 2020b), positive emotion words (Preoţiuc-Pietro et al., 2017; Slatcher et al., 2007) and a focus on the past (Robinson et al., 2015; Rodriguez, 2020b). Conversely, progressives are more inclined to use language related to anxiety (Preoţiuc-Pietro et al., 2017), affect (Rodriguez, 2020b), negative emotional words
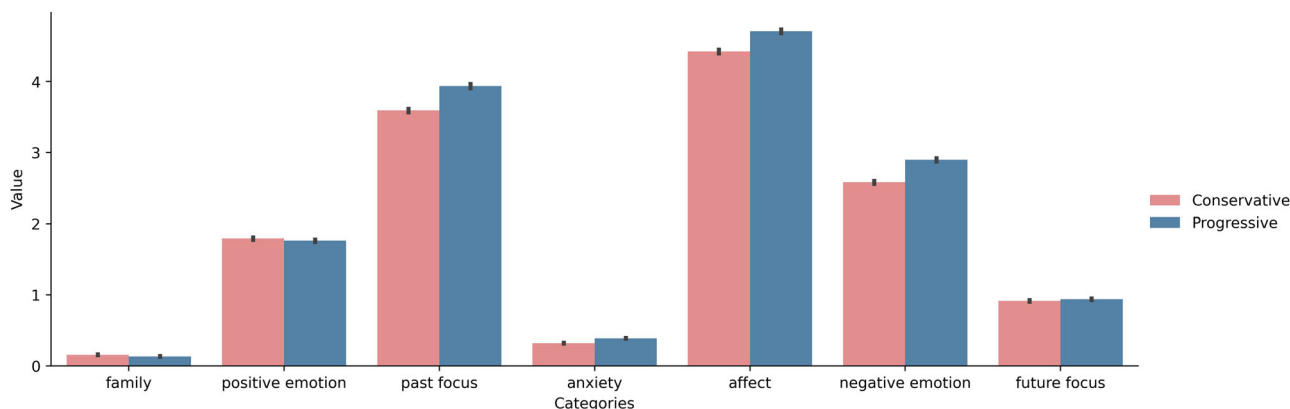
Fig. 1 Comparison of seven LIWC categories between conservative and progressive media. The bar chart shows the mean values for seven LIWC categories (Family, Positive Emotion, Past Orientation, Anxiety, Affect, Negative Emotion, and Future Orientation) in conservative outlets (blue bars) versus progressive outlets (red bars).

Table 2 T-test results comparing conservative and progressive media.

Value	Family	Positive Emotion	Past Focus	Anxiety	Affect	Negative Emotion	Future Focus
T-value	2.4351	−1.2553	−8.4220	−7.7478	−7.5169	−10.1455	−1.3981
p value	0.0149	0.2094	<0.0001	<0.0001	0.0013	0.0037	0.1159

Table 3 Results of model evaluation.

Model	Label	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
XGB	*Conservative*	**95.98**	89.88	92.83	93.06
	Progressive	90.49	**96.24**	93.27	
BERT	*Conservative*	93.06	**97.03**	**95.00**	**94.88**
	Progressive	**96.87**	92.71	**94.74**	

Values that are comparatively higher have been set in bold.

(Slatcher et al., 2007), and focus on the future (Robinson et al., 2015; Rodriguez, 2020b). Building on insights from these previous studies, the researchers constructed LIWC categories. Figure 1 presents the categories and comparison results based on these linguistic features and their associations with political orientations.

In the conservative group, the results showed the following averages and standard deviations for the LIWC categories: family ($M = 0.15$, SD = 0.40), positive emotion ($M = 1.79$, SD = 1.04), past focus ($M = 3.59$, SD = 1.68), anxiety ($M = 0.32$, SD = 0.36), affect ($M = 4.42$, SD = 1.71), negative emotion ($M = 2.58$, SD = 1.38), and future focus ($M = 0.91$, SD = 0.70).

For the progressive group, the results were as follows: family ($M = 0.13$, SD = 0.41), positive emotions ($M = 1.76$, SD = 1.10), past focus ($M=3.93$, SD=1.97), anxiety ($M = 0.38$, SD = 0.43), affect ($M = 4.70$, SD=1.68), negative emotions ($M = 2.89$, SD = 1.41), and future focus ($M = 0.93$, SD = 0.76). In comparison to the findings of previous studies, this study found similar relationships with previous findings. Specifically, conservative individuals showed higher average values than progressives in the categories of family and positive emotions, consistent with earlier research. On the other hand, progressives exhibited higher values for anxiety, affect, negative emotions, and future focus, consistent with the results of previous studies. However, one notable difference was observed in past focus, where progressives demonstrated higher values than conservative individuals, differing from previous findings.

To conduct a more rigorous verification of the statistically significant differences, we performed a *t*-test comparison, and the results are presented in Table 2. The analysis confirms that conservative media use more family-related expressions, while progressive media employ significantly more emotional expressions, anxiety-related terms, and negative expressions. However, unlike previous studies, our findings indicate that progressive media use past-focused expressions more frequently, whereas no significant differences were observed in positive emotion and future-oriented expressions.

Construction of classification model with media text. In this section, we describe the construction of the classification model and assess its performance. The dataset was divided into an 8:2 ratio for training and evaluation. The evaluation results are listed in Table 3 and Fig. 2. Based on the model evaluation, the BERT model outperformed the XGB model in almost all performance metrics. Thus, our results indicate that the BERT model is most suitable for this task.

Government text classification. Before performing classification with the models, we attempted to cluster vectorized documents using the *K*-means algorithm (Lloyd, 1982). *K*-means is a clustering algorithm widely used in data science, aiming to identify cluster structures characterized by high similarity within the same cluster and significant dissimilarity between different clusters
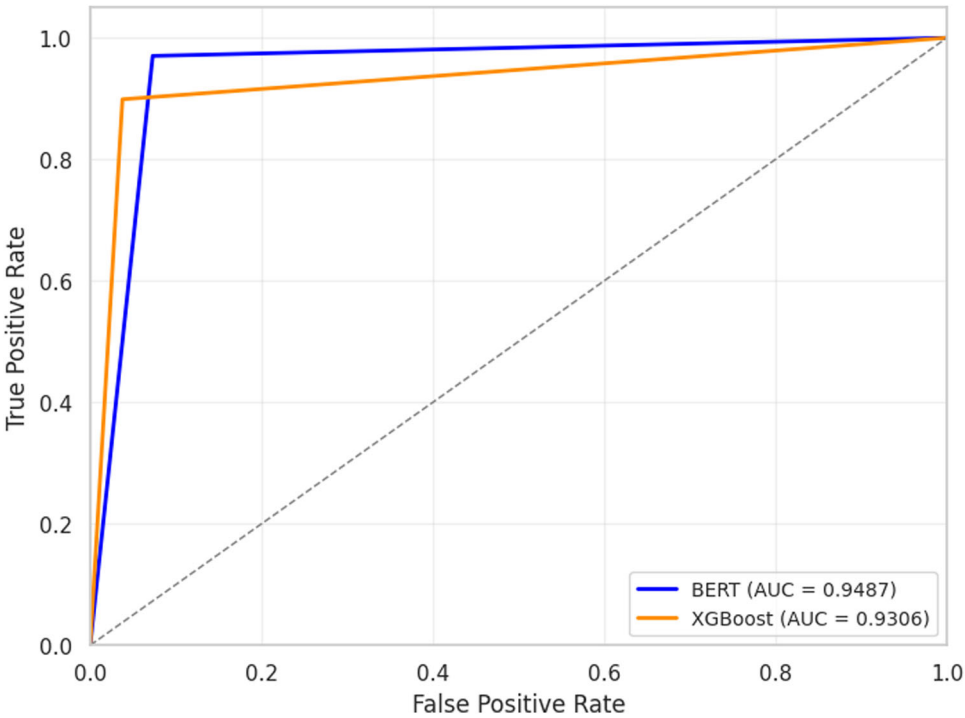
Fig. 2 ROC-AUC curves for BERT and XGBoost classifiers. ROC curves obtained on the held-out test set for a fine-tuned BERT model and an XGBoost model, demonstrating superior classification performance by BERT.
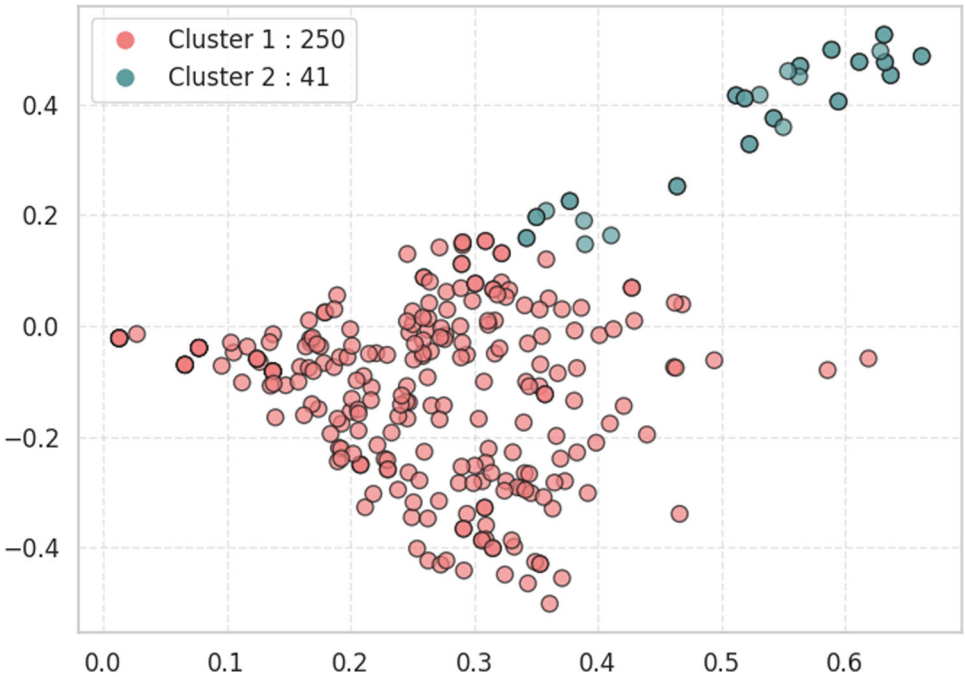

Fig. 3 K-means clustering of official U.S. government documents using BERT embeddings. Scatterplot of document representations encoded with BERT, reduced to two dimensions via PCA, and clustered with the K-means algorithm. Each point represents a single document, colored by its assigned cluster.

within a dataset (Sinaga and Yang, 2020). The clustering results are shown in Fig. 3, where we observed a skewed distribution of the data points. Subsequently, in Table 3, we confirmed that the BERT model achieved the best performance among the models considered.

Based on these results, we conducted a classification using the BERT model. The classification revealed a biased outcome, with 64 classified as conservative (21.99%) and 227 as progressive

(78%), which is consistent with the clustering results. Upon detailed scrutiny, all institutions exhibited a bias toward the progressive side, with White House showing the most progressive outcome at 88.23%. The classification results are listed in Table 4. Similar to the clustering results, the classification results display a skewed distribution. According to the model, we observed a tendency for government documents to be classified in a progressive orientation.

Table 4 Classification result of governments' official documents.

Label	White House	Treasury	Defense	State	Total
Conservative	14 (11.76%)	23 (34.32%)	20 (30.76%)	7 (17.50%)	**64 (21.99%)**
Progressive	105 (88.23%)	44 (65.67%)	45 (69.23%)	33 (82.50%)	**227 (78.00%)**

The overall "Total" results are highlighted in bold.

Table 5 Comparison of JSD Values with Mean and t-test.

JSD with Official Documents	1st	2nd	3rd	4th	5th	Mean	T value	P value
Progressive	.6777	.6796	.6795	.6756	.6809	.6787	2.7986	0.0468
Conservative	**.6593**	**.6653**	**.6782**	**.6603**	**.6377**	**.6602**		

Boldface is used to denote the lower JSD values.

Topic modeling

Verifictation of similarity between government and media text. In the previous section, we classified government documents by using a trained classification model. In this section, we compare the similarities between government documents and texts produced by progressive and conservative media through topic modeling. This allowed us to verify whether the results aligned with those of the classification model.

To measure similarity, we used the Jensen-Shannon divergence (JSD), which quantifies the similarity between two probability distributions. JSD is widely used to compare word distributions across different corpora or documents, helping analyze differences in themes and writing styles (Lin, 1991).

The experimental results (Table 5) show that the average JSD distance over five trials between government documents and progressive media was 0.6787, whereas the distance from conservative media was 0.6602. Statistical validation further confirms that Official Documents exhibit a textual distribution that is more similar to that of conservative media.

Verification of inherent intention in government text. This study aims to analyze the key opinions and core topics present in official documents issued by the U.S. government. To this end, we first examined the classification of documents and similarities in topic distributions; however, no distinct political inclinations were identified.

Accordingly, we employed BERTopic to extract the predominant topics in the government documents and identify the most salient terms within each topic. Our analysis revealed that out of a total of 291 documents, 279 were classified into specific topics, with key terms such as "Russian," "state," "United," "Russia," "company," "today," "president," "also," "designated," and "sanction" emerging as the most significant. These findings suggest that U.S. government documents predominantly address issues related to international relations, economic sanctions, government policies, and administrative measures. Notably, the frequent recurrence of terms like "Russian" and "Russia" indicates a strong connection to geopolitical developments, while the prominence of words such as "sanction" and "designated" underscores the substantial emphasis placed on discussions regarding economic and political measures.

Through this analysis, we systematically identified recurring concepts and their interrelations within U.S. government documents. This study provides valuable insights that may facilitate a more effective exploration of key issues in official government discourse and contribute to informed decision making in policy formulation.

Conclusions

In this study, our main objective was to assess the political orientation of news media concerning the Russia-Ukraine conflict. To achieve this, we collected news articles related to the Russia-Ukraine war through media outlets representing conservative and progressive perspectives in the United States, and selected and compared LIWC categories, taking into account the linguistic features previously observed in conservative and progressive contexts based on earlier research. The comparative results confirmed that news media articles produced based on the Russia-Ukraine conflict incorporated their unique political orientations and addressed RQ1 in this study.

Following a comparative analysis of articles from both perspectives, we trained the ML/DL models on a politically labeled dataset. The BERT model outperformed the conventional ML approaches (94.88% accuracy). Subsequent BERTopic analysis and JSD-based similarity measurement between the classification and topic models produced incongruent results. Examination of the dominant topics uncovered intentional framing around economic/political sanctions, offering partial but substantive insights for addressing RQ2.

Governments communicate their decisions and intentions both domestically and internationally through formalized documents. As texts are typically coherent and consistent to fulfill the producer's intentions (De Beaugrande and Dressler, 1981), in this experiment, we utilized the constructed model to examine the U.S. government's intentions concerning the Russia-Ukraine conflict. The analysis suggests that the nature of the U.S. government's official documents regarding the Russia-Ukraine war is a result of strategic framing aimed at fostering international solidarity and garnering domestic and global support. This underscores the need for further in-depth research on how government discourse influences the formation of public opinion. This approach provided valuable information about the government's stance and policies regarding the Russia-Ukraine conflict.

Although this study makes significant academic contributions, it has several limitations. First, news media coverage of the Russia-Ukraine conflict may include fact-based informational content, without explicit political nuances. The analysis of political orientation in neutral articles faces methodological constraints.

Second, the constructed dataset has methodological limitations, as it merely categorizes content into progressive or conservative classifications, without reflecting gradations in political learning. When texts exhibit similar probability scores for both conservative and progressive orientations, the model's forced binary classification may overlook nuanced complexities and subtle variations in political stance.

Therefore, future research should develop a multidimensional analytical framework capable of quantifying bias intensity in texts. Such an approach would enable a more comprehensive understanding of media representation patterns and the underlying political implications in the coverage of the Russia-Ukraine conflict.

Ethical and Societal Considerations

This study acknowledges potential ethical concerns in categorizing official government documents as conservative or progressive. While political orientation analysis is useful for understanding the ideological elements embedded in texts, excessive simplification or misuse may exacerbate political polarization. Therefore, the findings of this study should be interpreted responsibly with caution against their misuse as tools for deepening political divisions.

Data availability

All datasets in this study are available at https://github.com/dxlabskku/Russia-Ukraine-conflict.

Notes

1 https://www.pewresearch.org/
2 https://mediabiasfactcheck.com/

References

Abe JAA (2018) Personality and political preferences: the 2016 us presidential election. J Res Personal 77:70–82

Agrawal A, Fu W, Menzies T (2018) What is wrong with topic modeling? and how to fix it using search-based software engineering. Inf Softw Technol 98:74–88

Altmann GT (1998) The ascent of Babel: an exploration of language, mind, and understanding. OUP Oxford

Alturayeif N, Luqman H, Ahmed M (2023) A systematic review of machine learning techniques for stance detection and its applications. Neural Comput Appl 35:5113–5144

Baly R, Martino GDS, Glass J, Nakov P (2020) We can detect your bias: predicting the political ideology of news articles. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. ACL, Stroudsburg, PA, USA, p 4982–4991

Biesbroek R, Badloe S, Athanasiadis IN (2020) Machine learning for research on climate change adaptation policy integration: an exploratory UK case study. Reg Environ Change 20:85

Bisong E (2019) Building machine learning and deep learning models on Google cloud platform. In: a comprehensive guide for beginners. Apress, Berkeley, CA, p 59–64

Blanchard MA (2013) History of the mass media in the United States: an encyclopedia. Routledge, New York

Brunk IW, Hakimi M (2022) Russia, Ukraine, and the future world order. Am J Int Law 116:687–697

Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: 22nd ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, NY, p 785–794

Chen X, Madiyeva G (2022) The influence of political environment on language and the development of the modern Russia. J Ethn Cult Stud 9:212–225

Conover MD, Gonçalves B, Ratkiewicz J, Flammini A, Menczer F (2011) Predicting the political alignment of Twitter users. In: IEEE 3rd international conference on privacy, security, risk and trust and 2011 IEEE 3rd international conference on social computing. IEEE, Boston, MA, USA, p 192–199

De Beaugrande R-A, Dressler WU (1981) Introduction to text linguistics, vol. 1. Longman, London

Devatine N, Muller P, Braud C (2022) Predicting political orientation in news with latent discourse structure to improve bias understanding. In: 3rd workshop on computational approaches to discourse. association for computational linguistics, Gyeongju, South Korea, p 77–85

Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies. ACL, Minneapolis, MN, USA, p 4171–4186

Entman RM (1993) Framing: toward clarification of a fractured paradigm. J Commun 43:51–58

Entman RM (2007) Framing bias: media in the distribution of power. J Commun 57:163–173

Geis ML (2012)The language of politics. Springer Science & Business Media, Singapore

Graham J, Haidt J, Nosek BA (2009) Liberals and conservatives rely on different sets of moral foundations. J Personal Soc Psychol 96:1029

Grootendorst M (2022) Bertopic: neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv: https://arxiv.org/abs/2203.05794

Haidt J, Graham J (2007) When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. Soc justice Res 20:98–116

Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag process 5:1

Jones KL, Noorbaloochi S, Jost JT, Bonneau R, Nagler J, Tucker JA (2018) Liberal and conservative values: What we can learn from congressional tweets. Political Psychol 39:423–443

Kangas SE (2014) What can software tell us about political candidates? A critical analysis of a computerized method for political discourse. J Lang Polit 13:77–97

Kaplan A, Haenlein M (2019) Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 62:15–25

Kherwa P, Bansal P (2020) Topic modeling: a comprehensive review. EAI Endorsed Trans Scalable Inf Syst 7:e2

Kim MY, Johnson K (2022) Close: contrastive learning of subframe embeddings for political bias classification of news media. In: 29th international conference on computational linguistics. International Committee on Computational Linguistics, Gyeongju, Republic of Korea, p 2780–2793

Kirby P (2022) Why is Russia invading Ukraine and what does Putin want? vol 9. BBC News Retrieved. https://www.ianfeinhandler.com/iaclub/articles/Why_is_Russia_invading_Ukraine.pdf

Ko S, Cha J, Park E (2023) A machine learning approach to predict timely ferry services using integrated meteorological datasets. In: institution of civil engineers-transport. Emerald Publishing Limited, p 1–17

Körner R, Overbeck JR, Körner E, Schütz A (2022) How the linguistic styles of Donald Trump and Joe Biden reflect different forms of power. J Lang Soc Psychol 41:631–658

Kudelia S (2022) The Ukrainian state under russian aggression: resilience and resistance. Curr Hist 121:251–257

Külz J, Spitz A, Abu-Akel A, Günnemann S, West R (2023) United States politicians' tone became more negative with 2016 primary campaigns. Sci Rep. 13:10495

Lakoff G (2010) Moral politics: how liberals and conservatives think. University of Chicago Press, Chicago

Lee S, Kim J, Kim D, Kim KJ, Park E (2023) Computational approaches to developing the implicit media bias dataset: assessing political orientations of nonpolitical news articles. Appl Math Comput 458:128219

Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory 37:145–151

Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inf Theory 28:129–137

Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5:1093–1113

Ogunleye A, Wang Q-G (2019) Xgboost model for chronic kidney disease diagnosis. IEEE/ACM Trans Comput Biol Bioinform 17:2131–2140

Park S, Ko M, Kim J, Liu Y, Song J (2011) The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In: CSCW '11. Association for Computing Machinery, NY, p 113–122

Pennebaker JW, Boyd RL, Jordan K, Blackburn K (2015) The development and psychometric properties of LIWC2015. The University of Texas at Austin

Pew Research Center (2023) As Russian invasion nears one-year mark, partisans grow further apart on U.S. support for Ukraine. https://www.pewresearch.org/short-reads/2023/01/31/as-russian-invasion-nears-one-year-mark-partisans-grow-further-apart-on-u-s-support-for-ukraine/. Accessed 07 May 2023

Potrafke N (2018) Government ideology and economic policy-making in the United States–a survey. Public Choice 174:145–207

Preoţiuc-Pietro D, Liu Y, Hopkins D, Ungar L (2017) Beyond binary labels: Political ideology prediction of Twitter users. In: 55th annual meeting of the association for computational linguistics, Association for Computational Linguistics, Canada, vol 1, 729–740

Robinson MD, Cassidy DM, Boyd RL, Fetterman AK (2015) The politics of time: conservatives differentially reference the past and liberals differentially reference the future. J Appl Soc Psychol 45:391–399

Rodriguez J (2020a) A LIWC analysis of moral rhetoric in campaign memoirs by US Presidential candidates: 2000–2016. Ph.D. thesis, University of South Dakota

Rodriguez J (2020b) A LIWC analysis of moral rhetoric in campaign memoirs by US Presidential candidates: 2000–2016. Ph.D. thesis, University of South Dakota

Sinaga KP, Yang M-S (2020) Unsupervised k-means clustering algorithm. IEEE access 8:80716–80727

Slatcher RB, Chung CK, Pennebaker JW, Stone LD (2007) Winning words: Individual differences in linguistic style among us presidential and vice presidential candidates. J Res Personal 41:63–75

Sterling J, Jost JT, Bonneau R (2020) Political psycholinguistics: a comprehensive analysis of the language habits of liberal and conservative social media users. J Pers Soc Psychol 118:805

Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 29:24–54

Vajriyati S, Basuki LW, Lessy AK, Anieda KI, Kuswoyo LC, Meristiana M (2022) The effect of the Russia-Ukraine conflict on the potential use of nuclear weapons. J Soc Polit Sci 3:235–267

Walker SG (2009) The psychology of presidential decision making. Oxford University Press, Oxford

Wang W, Zhou H, He K, Hopcroft JE (2017) Learning latent topics from the word co-occurrence network. In: Du D, Li L, Zhu E, He K (eds) Theoretical computer science. NCTCS 2017. Communications in computer and information science, vol 768. Springer, Singapore

Whitehouse. Fact sheet: United States and G7 partners impose severe costs for Putin's war against Ukraine (2022). https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/08/fact-sheet-united-states-and-g7-partners-impose-severe-costs-for-putins-war-against-ukraine/. Accessed 11 May 2023

Whitehouse. Remarks by president biden on russia's unprovoked and unjustified attack on ukraine (2022). https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/02/24/remarks-by-president-biden-on-russias-unprovoked-and-unjustified-attack-on-ukraine/. Accessed 11 May 2023

Whitehouse. Fact sheet: on one year anniversary of Russia's invasion of Ukraine, Biden administration announces actions to support Ukraine and hold Russia accountable (2023). https://www.whitehouse.gov/briefing-room/statements-releases/2023/02/24/fact-sheet-on-one-year-anniversary-of-russias-invasion-of-ukraine-biden-administration-announces-actions-to-support-ukraine-and-hold-russia-accountable/. Accessed 11 May 2023

Acknowledgements

Author contributions

SeongKyu Ko: Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft. Junyeop Cha: Conceptualization, Validation, Writing - review & editing. Eunil Park: Validation, Writing - review & editing, Supervision, Project administration, Funding acquisition. Angel P. del Pobil: Writing - review & editing, Supervision, Project administration, Funding acquisition.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants

Informed consent

No human participants were involved in this study.Consent for publication Not applicable.

Additional information

Correspondence and requests for materials should be addressed to Junyeop Cha or Eunil Park.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.